CAPSTONE PROJECT

Introduction

Madrid is the capital of Spain. Only the city has 3.2 million inhabitants to which its needed to add another 3 million people who lives inside the surroundings of the city. It is a very touristy are, which is visited by near 8 million people every year. The great amount of population added to the number of visitors from every country of the world makes a Madrid a multicultural city where coexist wide range of restaurants.

Business problem

The city contains hundreds of restaurants so the inauguration of a new one need to be a carefully decision taken because choosing the right location for it can make the difference between a huge success or a big failure in the business. The main goal of this project is to analyse the current restaurants in the 21 boroughs that form the city of Madrid in order to decide a suitable location and the most convenient type of restaurant.

Target audience

The target audience of this project are possible stakeholders or investors that has in their plans to open a new restaurant in Madrid. It also can be used for current restaurants to evaluate if their decision has been appropriate or it would be more convenient to tweak their type of cuisine.

Data description

The used data consists of:

- Location of Madrid boroughs.
- List of all the restaurants in the different Madrid boroughs using the Foursquare API

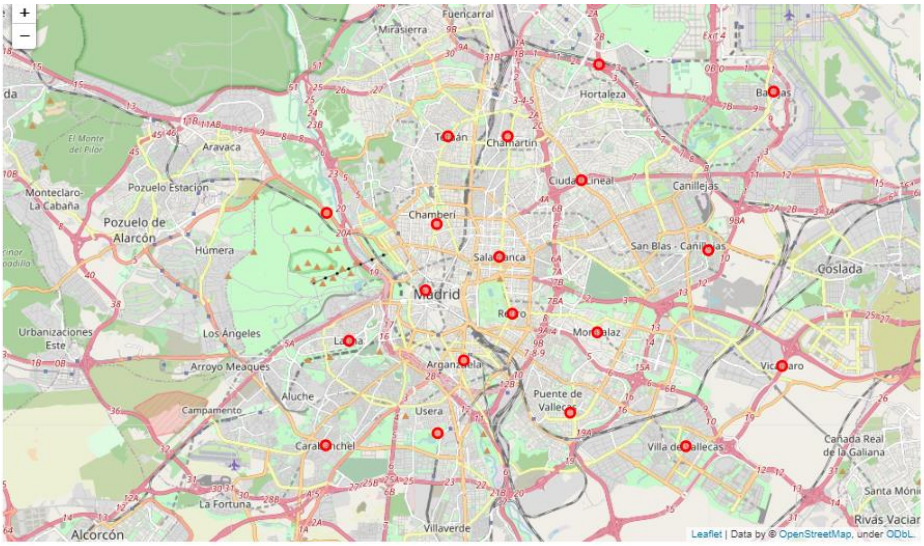
The procedure will be the following: First, it will be obtained the location of the different boroughs of Madrid. After that, they will be plotted in a map to study its disposition. Next, it will be obtained the venues in the boroughs using a radius of 500. This measurement has been chosen to avoid duplicates due the proximity and the small area of some of the boroughs. Then using K-Means clustering the different restaurants will be classified and plotted using the Folium library. With this visualization it will be possible to study and determine which can be the best location for a certain restaurant type.

Data acquisition source

On the one hand, the location of Madrid boroughs will be obtained by scrapping in the following website: https://es.wikipedia.org/wiki/Anexo:Distritos_de_Madrid

After pre-processing the data and adding the location using geopy library, the dataframe looks like:

The location of the boroughs in the Madrid map is the following:



Data Science Methodology

First, the data scrapped of the Madrid boroughs is cleaned to keep the values that are used to use the Foursquare API. Next, for each borough it has been used the Foursquare API to find the principal venues:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro	40.417653	-3.707914	Plaza de Isabel II	40.418114	-3.709397	Plaza
1	Centro	40.417653	-3.707914	TOC Hostel	40.417264	-3.705928	Hostel
2	Centro	40.417653	-3.707914	Plaza Mayor	40.415527	-3.707506	Plaza
3	Centro	40.417653	-3.707914	Gyoza Go!	40.416179	-3.708612	Dumpling Restaurant
4	Centro	40.417653	-3.707914	Palacio de Gaviria	40.417139	-3.706044	Art Museum
...	...	...	...	...	...	...	...
554	Barajas	40.473318	-3.579845	Mercadillo Barajas	40.470179	-3.577668	Flea Market
555	Barajas	40.473318	-3.579845	Metro Barajas	40.475768	-3.582535	Metro Station
556	Barajas	40.473318	-3.579845	Burger King	40.473312	-3.582063	Fast Food Restaurant
557	Barajas	40.473318	-3.579845	Tryp Alameda Aeropuerto	40.469134	-3.580105	Hotel
558	Barajas	40.473318	-3.579845	Espléndido Breakfast Buffet	40.469837	-3.582108	Breakfast Spot

559 rows × 7 columns

Next, the dataframe obtained has been filtered to keep the venues related with restaurants and then ordered by venue category:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
154	Salamanca	40.427045	-3.680602	Dingo	40.427247	-3.684072	American Restaurant
313	Chamberi	40.436247	-3.703830	Rochela Restaurante	40.433575	-3.702525	American Restaurant
531	Barajas	40.473318	-3.579845	La Torino	40.473706	-3.578169	Argentinian Restaurant
517	Barajas	40.473318	-3.579845	Finca Lucero	40.470693	-3.583052	Argentinian Restaurant
452	Ciudad Lineal	40.448431	-3.650495	La Vaca Argentina	40.451483	-3.649788	Argentinian Restaurant
...	...	...	...	...	...	...	...
497	Villa de Vallecas	40.373958	-3.612163	Nueva York	40.370319	-3.612369	Tapas Restaurant
33	Centro	40.417653	-3.707914	El Mollete	40.419913	-3.710503	Tapas Restaurant
207	Salamanca	40.427045	-3.680602	Thaidy	40.423578	-3.679922	Thai Restaurant
546	Barajas	40.473318	-3.579845	Barajas Doner Kebab	40.472209	-3.579033	Turkish Restaurant
290	Tetuan	40.460821	-3.699520	Mr. Vu	40.465107	-3.697740	Vietnamese Restaurant

195 rows × 7 columns

It has been obtained a total of 28 kinds of restaurant that are showed in the next list:

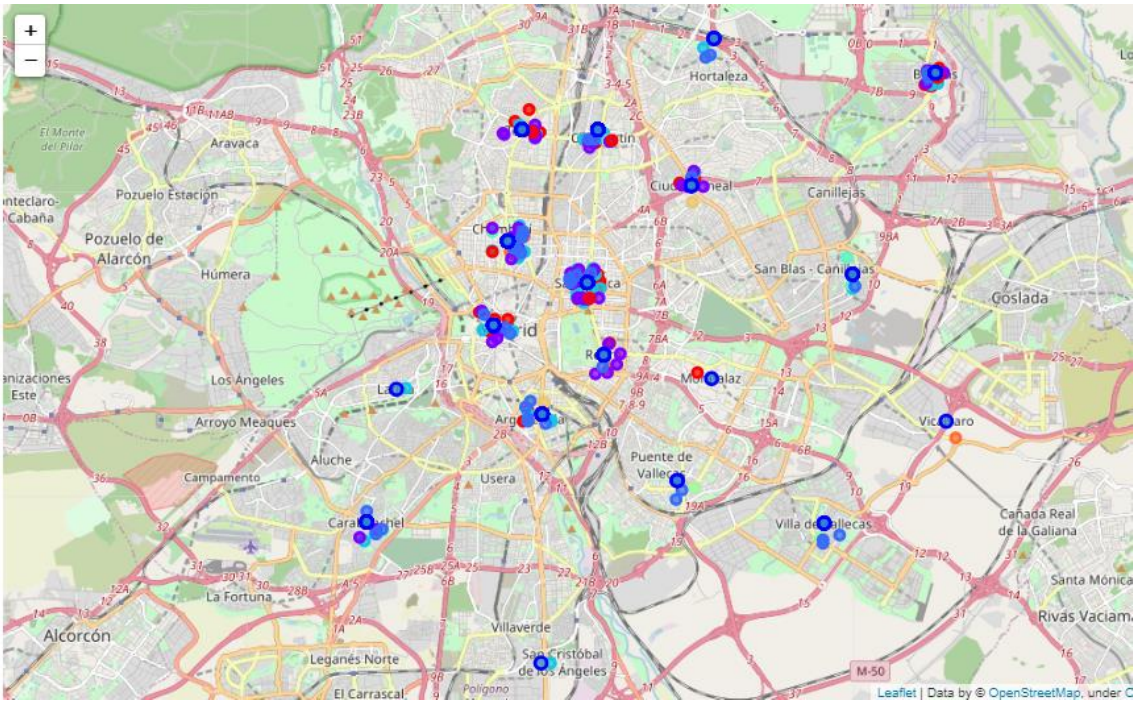
```
['American Restaurant',
 'Argentinian Restaurant',
 'Asian Restaurant',
 'Brazilian Restaurant',
 'Cajun / Creole Restaurant',
 'Chinese Restaurant',
 'Comfort Food Restaurant',
 'Dumpling Restaurant',
 'Falafel Restaurant',
 'Fast Food Restaurant',
 'French Restaurant',
 'Himalayan Restaurant',
 'Indian Restaurant',
 'Italian Restaurant',
 'Japanese Restaurant',
 'Korean Restaurant',
 'Mediterranean Restaurant',
 'Mexican Restaurant',
 'Middle Eastern Restaurant',
 'Ramen Restaurant',
 'Restaurant',
 'Seafood Restaurant',
 'Spanish Restaurant',
 'Sushi Restaurant',
 'Tapas Restaurant',
 'Thai Restaurant',
 'Turkish Restaurant',
 'Vietnamese Restaurant']
```

The next step has been to use one-hot encoding to create dummy variables for the type of restaurants as they are categorical variables.

	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dumpling Restaurant	Falafel Restaurant	Fast Food Restaurant	...	Middle Eastern Restaurant	Ramen Restaurant	Re:
154	1	0	0	0	0	0	0	0	0	0	...	0	0	
313	1	0	0	0	0	0	0	0	0	0	...	0	0	
531	0	1	0	0	0	0	0	0	0	0	...	0	0	
517	0	1	0	0	0	0	0	0	0	0	...	0	0	
452	0	1	0	0	0	0	0	0	0	0	...	0	0	

5 rows × 28 columns

The next step has been to use a Machine Learning algorithm to classify the restaurants. Concretely, it has been used the K-Means algorithm with 8 clusters. After that the classified restaurants has been plotted in the Madrid map using its location:



Results

The clusters obtained are shown in the following table:

Cluster	Restaurant Type
0 (Red)	Italian-Argentinian-Indian-American Restaurants
1 (Purple)	Spanish Restaurants
2 (Darkblue)	Tapas Restaurants
3 (Clearblue)	Generic Restaurants
4 (Cyan)	Seafood Restaurants
5 (Green)	Asian Restaurants
6 (Yellow)	Mexican Restaurants
7 (Orange)	Mediterranean Restaurant

Observations and Discussions

1. The Centro borough which is the most touristic concentrate a high density of Spanish and Tapas Restaurant where typical Spanish meals are provided.
2. In general, taking in account all the boroughs, the blue-purple colour is the predominant, which means that typical Spanish restaurant are the most extended.
3. If we continue focusing in the most centrical boroughs the most interesting restaurant possibilities are international restaurants such as Mexican or Italian due the lower amount of establishment focused on that foods.
4. In areas like Chamartin or Tetuan it can be observed a lack of Asian restaurant which means that opening one of this establishments can be good decision based also in the population of these areas.
5. Due to the high amount of borough, quite considerations can be extracted attending to the classification done, with just looking the plotted map.

Conclusion

With the conducting of the project it has been realised a study of the different restaurants in the boroughs in the city of Madrid. This study can help possible stakeholders that are interested in starting a new restaurant business in Madrid.